DOCUMENT RESUME

ED 467 381 TM 034 307

AUTHOR	Henson, Robin K.
TITLE	The Logic and Interpretation of Structure Coefficients in Multivariate General Linear Model Analyses.
PUB DATE	2002-04-03
NOTE	35p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Heuristics; *Multivariate Analysis
IDENTIFIERS	*General Linear Model; *Structure Coefficients

ABSTRACT

In General Linear Model (GLM) analyses, it is important to interpret structure coefficients, along with standardized weights, when evaluating variable contribution to observed effects. Although often used in canonical correlation analysis, structure coefficients are less frequently used in multiple regression and several other multivariate analyses. This paper discusses and demonstrates the role of structure coefficients in multivariate analyses by: (1) illustrating structure coefficients in the univariate context with multiple regression; and (2) using canonical correlation analysis to demonstrate structure coefficients in the multivariate context. A small heuristic data set is used to make the demonstration concretely accessible for applied researchers. (Contains 4 tables, 3 figures, and 22 references.) (Author/SLD)

struct_mult.doc

Running head: STRUCTURE COEFFFICIENTS

The Logic and Interpretation of Structure Coefficients in Multivariate General Linear Model Analyses

Robin K. Henson

University of North Texas

Paper presented at the annual meeting of the American Educational Research Association, April 3, 2002, New Orleans.

Correspondence concerning this paper should be sent to rhenson@unt.edu.

struct_mult.doc

Abstract

In General Linear Model (GLM) analyses, it is important to interpret structure coefficients, alongside standardized weights, when evaluating variable contribution to observed effects. Although often used in canonical correlation analysis, structure coefficients are less frequently employed in multiple regression and several other multivariate analyses. The present paper discusses and demonstrates the role of structure coefficients in multivariate analyses by (a) illustrating structure coefficients in the univariate context with multiple regression and (b) using canonical correlation analysis to demonstrate structure coefficients in the multivariate context. A small heuristic data set is used to make the demonstration concretely accessible for applied researchers.

struct_mult.doc

The Logic and Interpretation of Structure Coefficients in Multivariate General Linear Model Analyses

It is commonly known that the General Linear Model (GLM) serves as a general analytic system guiding all classical parametric analyses. Cohen (1968) demonstrated multiple regression as the univariate GLM. Knapp (1978) later illustrated that canonical correlation subsumed not only multiple regression but other multivariate analyses as the multivariate GLM umbrella. Structural equation modeling has since been shown as the most general case of the GLM, allowing simultaneous measurement and substantive modeling as part of the same analysis (Bagozzi, Fornell, & Larcker, 1981; Fan, 1997).

Understanding the foundational components of the GLM affords researchers wider utility and application of the various GLM analyses. Importantly, all GLM analyses have certain analytic characteristics in common. All analyses (a) are correlational in nature, (b) invoke a system of weights that are applied to observed variables to create synthetic (i.e., latent or unobserved) variables, (c) typically focus on the synthetic variables for analytic interest, and (d) yield r^2-type effect sizes.

Determining Variable Importance

In applied research, it is often important to identify variables that contribute to the model being tested. For example, an educational psychologist may use multiple regression to evaluate whether self-esteem, self-concept, and self-efficacy are predictive of academic achievement. The researcher likely cares about which, if any, of the variables is able to predict achievement and to what degree. Identification of variable importance, then, is fundamental to many of the analyses we conduct.

However, within the GLM, all analyses yield r^2-type effect sizes that must be considered prior to evaluating what variables contributed to this effect. It makes no sense, for example, to have a miniscule (and uninterpretable) effect size and yet try to identify variables that contributed to that effect. Accordingly, Thompson (1997) articulated a two-stage hierarchal decision strategy that can be used to interpret any GLM analysis:

> All analyses are part of one general linear model. . . . When interpreting results in the context of this model, researchers should generally approach the analysis hierarchically, by asking two questions:
>
> Do I have anything? (Researchers decide this question by looking at come combination of statistical

significant tests, effect sizes . . . and replicability evidence.)

- If I have something, where do my effects originate? (Researchers often consult both the standardized weights implicit in all analyses and structure coefficients to decide this question.) (p. 31)

Once notable effects have been isolated, then (and only then) interpretation shifts to the identification of what variables in the model have contributed to that effect. Traditionally, the weights (often standardized) present in all GLM analyses are examined to judge the contribution of a variable to the effect observed. Using regression as an example, many researchers would discount the value of a variable with a small or near-zero β (beta) weight.

The sole interpretation of standardized weights, however, can lead to erroneous conclusions about variable importance. Burdenski (in press), Courville and Thompson (2001), Thompson and Borrello (1985) have documented the drawbacks of only consulting standardized weights in multiple regression. In GLM analyses, it is also important to interpret structure coefficients, alongside standardized weights, when evaluating variable contribution to the observed effect.

Structure coefficients are much less understood within the GLM as compared to the role of standardized weights.

Nevertheless, the reporting and interpretation of structure coefficients is critical to identification of variable importance in both univariate and multivariate analyses.

Purpose

The purpose of the present paper is to discuss and demonstrate the role of structure coefficients in multivariate analyses. Accordingly, this paper will (a) illustrate structure coefficients in the univariate context with multiple regression and (b) use canonical correlation analysis to demonstrate structure coefficients in the multivariate context. A small heuristic data set is used to make the demonstration concretely accessible for applied researchers.

Multiple Regression as a Univariate Example

Where Does an Effect Size Come From?

Fundamental to interpreting any GLM analysis is the size of the obtained effect, whether that effect be a variance-accounted-for (e.g., R^2, η^2) or mean difference (e.g., Cohen's *d*) statistic (cf. Henson & Smith, 2000; Snyder & Lawson, 1993; Wilkinson & APA Task Force on Statistical Inference, 1999). In regression, the effect size of interest is R^2, which (when multiplied by 100) is the percentage of variance in the dependent variable that can be explained by the predictor variables.

In a hypothetical scenario where two predictors are perfectly uncorrelated, the effect size is the sum of the squared correlations between each predictor and the dependent variable (Y):

$$R^2 = r_{YX1}^2 + r_{YX2}^2. \quad (1)$$

Equation (1) makes explicit the fact that the relationships between each predictor and Y are critical to obtaining the overall effect.

In the real world, however, predictors are usually correlated to some degree. In such cases the predictors may explain the same variance in Y, and use of equation (1) would be inappropriate because dual credit would be given to more than one predictor. Standardized weights (β) can be derived, however, that "split up" the shared variance among the predictors so no two predictors are given credit for the same explained variance in Y. The appropriate equation then becomes:

$$R^2 = \beta_1 r_{YX1} + \beta_1 r_{YX2}. \quad (2)$$

In the actual analysis, of course, the β weights are applied to the observed predictor scores (in Z score form) in a linear equation to yield a synthetic variable consisting of predicted $\hat{Y}$ scores that are as close as possible to the actual Y scores (for ordinary least squares regression):

$$\hat{Y} = \beta_1 X_1 + \beta_1 X_2. \quad (3)$$

Because we care about how close the synthetic $\hat{Y}$ scores are to the observed Y scores, the effect size can also be stated as the squared correlation between the predicted scores and the dependent variable:

$$R^2 = r_{Y\hat{Y}}^2. \tag{4}$$

Equation (4) informs us that the synthetic variable, $\hat{Y}$, is critical in the regression, and therefore critical in result interpretation.

Additionally, because the effect is in a squared metric, we can conceptualize the relationship between the predictors and Y graphically with Venn diagrams by representing the sum of squares of each variable. For example, in a multiple regression, assume the R^2 = .75 and the relationships between the two predictors and Y are r_{YX1}^2 = .50 and r_{YX1}^2 = .50. If we were to sum the individual predictors squared relationships with Y, we would get 100% explained variance, a result larger than the 75% effect! This tells us the two predictors must be explaining the some of the same part of the Y variance. If we assume the relationship between the two predictors is r_{X1X2}^2 = .25, then the graphical representation of the model might look like Figure 1.

INSERT FIGURE 1 ABOUT HERE

Figure 1 demonstrates a case where both predictors explain some of the Y variance (the $\hat{Y}$ area), but also explain some of the *same part* of the Y variance (the double slashed area). It is clear that both predictors are equally effective in predicting the dependent variable. However, the question for the linear equation (3) is: "What will be the magnitude of the β weights?" Because standardized weights cannot allow dual credit be assigned to more than one variable for predicted ($\hat{Y}$) area, within any regression the β weights will be derived to either (a) arbitrarily "split up" the shared predicted area or (b) arbitrarily assign the entire portion to one of the variables.

However, should the shared area in Figure 1 be disproportionately divided between the predictors, then one β may be arbitrarily larger than the other, and therefore suggest that one variable is more important or contributes more to the predicted area than the other. Furthermore, if the study were to be conducted again, the βs may reverse their magnitudes for the two variables, a dilemma known as the "bouncing beta" problem.

What is the ''Structure'' of the Effect?

Because β weights cannot be examined to clearly identify the relationships between the predictors and the dependent variable when the predictors are correlated, more information is *necessarily* needed to interpret variable importance. Further, in the GLM, we almost always are concerned with the synthetic variables for interpretation purposes. In multiple regression, the $\hat{Y}$ predicted variable is our primary focus, as suggested by equation (4). In Figure 1, the sum of squares of $\hat{Y}$ is represented by the slashed area, which is 75% (R^2) of the total sum of squares of the dependent variable (or, put alternatively, $R^2 = SS_{explained}/SS_{total}$). The explanation of what variables contributed to this effect is central to interpreting variable importance.

Structure coefficients are called such because they inform us as to the "structure" or makeup of the effect represented by the synthetic variable $\hat{Y}$. By definition, a structure coefficient is a simple bivariate correlation between an observed variable (e.g., predictor) and a synthetic variable (e.g., $\hat{Y}$). Notice that because they are bivariate correlations, structure coefficients do not take into account the collinearity between the predictors, and therefore shed important light on the importance of predictors.

In the Figure 1 example, the *squared* structure coefficients between $\hat{Y}$ and X1 and X2 would both be .67, because both predictors can account for two-thirds of the explained effect ($\hat{Y}$ sum of squares) in and of themselves. Note that both structure coefficients (unsquared) would be the square root of .67, or +/-

.82, which is interpreted just as a Pearson *r*. If we sum the two squared structure coefficients (.67 + .67 = 1.34) the result is larger than 1.00, due to the fact that there is shared explained area. It should be apparent that if the predictors were perfectly uncorrelated, then the sum of the squared structure coefficients would be 1.00, because the predictors would account for unique portions of the $\hat{Y}$ variance.

If standardized weights inform the researcher what variables are getting credit for the effect, then structure coefficients inform the researcher what variables could have gotten credit for the effect. Both coefficients are important, and both coefficients should be reported and interpreted in published research. In the above example, the claim that one variable is better than the other would be unfounded (yet definitely possible if consulting βs), and examination of structure coefficients points to equal contributory value.

So What's the Problem with Multicollinearity?

Multicollinearity, or the presence of correlation between predictors, is often cited as a problem in multiple regression and therefore to be avoided. Stevens (2002), for example, stated:

> Multicollinearity poses a real problem for the researcher using multiple regression for three reasons:
>
> 1. It severely limits the size of *R*, because the

predictors are going after much of the same variance on *y*. . . .

2. Multicollinearity makes determining the importance of a given predictor difficult because the effects of the predictors are confounded due to the correlations among them.

3. Multicollinearity increases the variances of the regression coefficients. (p. 91-92)

These concerns are merited if considering only β weights, but they largely become mute when interpreting structure coefficients. Regarding (1), the predictors may explain the same part of Y variance, but R is not artificially limited. Of course, R will not get bigger unless additional portions of dependent variable variance are explained, but within the GLM, the addition of predictor variables will only result in R^2 either remaining the same (no additional variance explained) or getting larger. Regarding (2), structure coefficients clarify variable importance as noted above. Regarding (3), decisions based solely on β weights may be impacted by inflated standard errors if using statistical significance tests. However, structure coefficients are not impacted by inflated standard errors as they are descriptive correlational measures. Of course, β standard errors are only relevant when the researcher depends on statistical significance testing, which may obfuscate variable contributions due to the impact of sample size on null hypothesis tests. The bottom line is that multicollinearity is not a problem in multiple regression, and therefore not in any other GLM analysis, if the researcher invokes structure coefficients in addition to standardized weights. In fact, in some multivariate analyses, multicollinearity is actually encouraged, say, for example, when multi-operationalizing a dependent variable with several similar measures.

Canonical Correlation as a Multivariate Example

Canonical correlation analysis (CCA) is a natural multivariate extension of multiple regression with which researchers can examine the relationship between several predictors and several dependent variables simultaneously (Henson, 2000; Thompson, 1984, 1991). In CCA, the several predictors are linearly combined into one synthetic predictor variable. This process is directly analogous to the creation of $\hat{Y}$ in multiple regression. However, in CCA, the dependent variables are also linearly combined to create one synthetic criterion variable (also analogous to $\hat{Y}$). The canonical correlation itself is nothing more than a Pearson r correlation between the synthetic predictor and synthetic criterion variables. In CCA, this pair of synthetic variables is created for each canonical function (variate). The first function maximizes shared variance between the observed predictor and dependent variables. Subsequent functions are created (analogous to factors in factor analysis) that maximize explained variance for the residual (e.g., unexplained) variance left over from the previous function.

In CCA, then, there are two synthetic variables for each function. Because structure coefficients are the correlation between an observed and a synthetic variable, there are structure coefficients that explain relationships between both the predictor and criterion synthetic variables and their respective observed variable sets.

Table 1 presents a heuristic data set that will be used to help make the present discussion concrete. In this hypothetical example, the researcher is investigating whether two variables related to adult attachment, secure and preoccupied attachment styles, are predictive of variation in personality styles as measured by the ``Big Five'' factors: neuroticism, extraversion, openness, agreeableness, and conscientiousness. Data are presented as T scores for 10 people. For SECURE and PREOCC, Z scores are parenthetically presented for later use.

INSERT TABLE 1 ABOUT HERE

The CCA for SECURE and PREOCC predicting NEURO, EXTRA, OPEN, AGREE, and CONSC yielded a squared canonical correlation of .682 for the first function and .163 for the second function. (Note: There will be as many functions as there are variables in the smaller variable set, which in this case is two.) The Appendix presents the SPSS syntax used for the CCA. Supposing we deem only the first function's effect noteworthy, we now concern ourselves with defining this function by identifying what variables contributed to the effect.

Table 2 presents the standardized canonical function coefficients (directly analogous to β weights) and the structure coefficients for the first function. Examination of only the function coefficients might lead one to conclude the first function was largely the result of PREOCC predicting NEURO and AGREE. The prediction is positive due to the fact that all the structure coefficients have the same sign. Note as well that we cannot tell directionality by simply consulting the standardized weights, because when the observed variables are correlated, the standardized weights are not direct measures of relationship. In the present example, this is made explicit by the existence of a function coefficient greater than one for NEURO.

INSERT TABLE 2 ABOUT HERE

However, examination of the structure coefficients (r_s) and their squared values provides a more complete picture of the variable relationships. PREOCC is still clearly the best predictor, but it would be inappropriate to label SECURE as useless as it can account for over half of the synthetic predictor variable by itself, a fact obfuscated by the small standardized weight for SECURE. Additionally, if *only* consulting the function coefficients, it appears clear that NEURO and AGREE are the predominant criterion variables with the rest of the variables having low or near zero coefficients. Although NEURO is the dominant criterion variable, the squared structure coefficients inform us that EXTRA in fact can account for about one-third of the synthetic criterion variable while AGREE can account for less than 3%, in spite of having the second largest standardized weight!

Certainly consultation of only standardized weights can mask important relationships when variables are correlated. Also certainly, consultation of structure coefficients clarifies variable relationships in the presence of multicollinearity, which is almost always present in applied research.

Construction of the Synthetic/Latent Variables

Table 3 presents the calculations used to create the synthetic predictor variable for all 10 cases from Table 2. These calculations make explicit how the standardized weights are applied to the observed scores (in Z score form) to create the synthetic predictor variable, which as noted is directly analogous to $\hat{Y}$ in multiple regression. Remember that in the CCA, a similar equation is used that combines the dependent variables into one synthetic dependent variable. The present discussion, however, will focus only on the predictor side of the equation.

INSERT TABLE 3 ABOUT HERE

Canonical Structure Coefficients

The structure coefficients for SECURE and PREOCC can now be directly calculated from the Table 3 results as the correlations between the observed SECURE and PREOCC scores and the synthetic predictor variable created from those scores. These correlations are .752 and -.973 for SECURE and PREOCC, respectively. Note that these structure coefficients match exactly those in Table 2 created by the CCA.

Figures 2 and 3 graphically display the relationships between SECURE and PREOCC with the synthetic variable. In this scenario, both variables have strong relationships with the synthetic variable, and therefore should be considered in interpretation. Of course, consultation of only the function coefficients would not have led to the same conclusion.

INSERT FIGURES 2 and 3 ABOUT HERE

Structure Coefficients in Other Multivariate Analyses

Structure coefficients are present throughout the GLM, and typically are necessary for result interpretation. However, the literature is inconsistent in defining the role of structure coefficients in various GLM analyses. Burdenski (in press) and Courville and Thompson (2001) have documented that authors seldom report structure coefficients in multiple regression and almost exclusively only consult weights when determining variable importance.

In CCA, there is some consensus that structure coefficients are necessary (cf. Meredith, 1964; Thompson, 1984). As Levine (1977) argued:

> I specifically say that one has to do this [interpret structure coefficients] since I firmly believe as long as one wants information about the nature of the canonical correlation relationship, not merely the computation of the [synthetic variable] scores, one must have the structure matrix. (p. 20)

Cohen and Cohen (1983) also noted that "interpretation of a given canonical variate is best undertaken by means of the structure coefficients" (p. 456).

Because CCA is the multivariate GLM, and because structure coefficients are critical for CCA interpretation, it stands to reason that interpretation of other GLM analyses would also require structure coefficients. As Huberty (1994) explained,

> if a researcher is convinced that the use of structure *r*s makes sense in, say, a canonical correlation context, he or she would also advocate the use of structure *r*s in the contexts of multiple correlation, common factor analysis, and descriptive discriminant analysis. (p. 263)

However, like in multiple regression, structure coefficients are often ignored in other multivariate analyses. In factor analysis, for example, factors are uncorrelated when an orthogonal rotation is used (e.g., varimax). In such cases, the factor pattern matrix is the same as the factor structure matrix. The factor structure matrix is found by multiplying the factor pattern matrix ($\mathbf{P}_{V \times F}$) with the factor correlation matrix ($\mathbf{R}_{F \times F}$). When the factors are uncorrelated, $\mathbf{R}_{F \times F}$ is an identity matrix. Therefore, the structure matrix ($\mathbf{S}_{V \times F}$) will be the same as the pattern matrix, such that:

$$\mathbf{P}_{V \times F} \, \mathbf{R}_{F \times F} = \mathbf{S}_{V \times F}. \qquad (5)$$

This outcome is analogous to a regression with perfectly uncorrelated predictors, and interpretation of a separate structure matrix is unnecessary. However, when the factors are correlated via an oblique rotation, the pattern and structure matrices *will not be identical*, and both matrices should be reported and interpreted, just as one would interpret β weights and structure coefficients in regression in the presence of correlated predictors. Unfortunately, empirical reviews of exploratory factor analyses indicate that structure matrices are often ignored (Henson, Capraro, & Capraro, 2001; Henson & Roberts, in press).

Nevertheless, structure coefficients are present throughout the GLM and should be consulted when considering variable importance. Table 4 lists several multivariate analyses, along with common names for the standardized weights used in the analyses. Also given is a description of what structure coefficients are correlating in a given analysis.

INSERT TABLE 4 ABOUT HERE

Discussion

When observed predictors/variables are correlated and combined into a synthetic variable, the definition of observed effects (e.g., R^2) must invoke examination of both standardized weights and structure coefficients. This is true in multiple regression, canonical correlation analysis, and throughout the GLM.

One reasonable alternative to examining structure coefficients in multiple regression is to consult the correlations between the predictors and the dependent variable directly. Indeed, another way to derive structure coefficients is to divide the correlation between the predictor and dependent variable by the multiple R:

$$r_{XY} \ / \ R \ . \qquad (6)$$

This equation informs us that all regression structure coefficients are the zero-order correlations between predictors and the dependent variable divided by a constant (R). Therefore, these zero-order correlations contain the same information as the structure coefficients. However, the information is in a different metric, with structure coefficients representing relationship with the synthetic effect ($\hat{Y}$), which is of primary interest. So, the decision about whether to interpret the zero-order correlations with the dependent variable or structure coefficients depends on the researcher's desire to describe his or her results in terms of the effect obtained or the observed variables.

Pedhazur (1997) also argued that structure coefficients can be excessively large even when the effect size is not noteworthy. For example, a tiny R^2 = .01 effect might be found, and if this effect is almost entirely due to one of the predictors, the structure coefficient for that predictor might be, say, .90. However, within the hierarchal strategy for interpreting all GLM analyses, one would never interpret the origins of an effect without first declaring the effect to be worth interpreting. In this context, the concern about misinterpretation seems unwarranted.

Furthermore, within multivariate analyses, there is not a single dependent variable with which to correlate the observed predictors. In this context, structure coefficients are essential for interpretation.

The present paper has demonstrated the role of both standardized weights and structure coefficients in univariate and multivariate analyses. Both coefficients are important for determining variable importance. Unfortunately, structure coefficients are often ignored, and overdependence is placed on standardized weights, perhaps resulting in misinterpretation of substantive findings. The present paper may serve to inform applied researchers about (a) the presence of structure coefficients, (b) the conceptual underpinnings of what a structure coefficient is, and (c) how to interpret these coefficients to determine variable importance in the presence of multicollinearity.

References

Bagozzi, R.P., Fornell, C., & Larcker, D.F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research, 16*, 437-454.

Burdenski, T. (in press). The importance of structure coefficients in multiple regression: A review with examples from published literature. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 6). Stamford, CT: JAI Press.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*, 426-443.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement, 61*, 229-248.

Fan, X. (1997). Canonical correlation analysis and structural equation modeling: What do they have in common? *Structural Equation Modeling, 4*, 65-79.

Henson, R. K. (2000). Demystifying parametric analyses: Illustrating canonical correlation as the multivariate general linear model. *Multiple Linear Regression Viewpoints, 26(1)*, 11-19.

Henson, R. K., Capraro, R. M., & Capraro, M. M. (2001, November). *Exploratory factor analyses in published educational research.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR.

Henson, R. K., & Roberts, J. K. (in press). Exploratory factor analysis reporting practices in published research. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 6). Stamford, CT: JAI Press.

Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force report and current trends. *Journal of Research and Development in Education, 33*(4), 284-294.

Huberty, C. J (1984). *Applied discriminant analysis.* New York: Wiley and Sons.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin, 85*, 410-416.

Levine, M. S. (1977). *Canonical correlation and factor* *comparisons*. Beverly Hills, CA: Sage.

Meredith, W. (1964). Canonical correlations with fallible data. *Psychometrika, 29*, 55-65.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Forth Worth, TX: Harcourt Brace.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *The Journal of Experimental Education, 61*, 361-377.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Newbury Park, CA: Sage.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24*, 80-95.

Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. *Educational Researcher, 26*(5), 29-32.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement, 45*, 203-209.

Wilkinson, L., & American Psychological Association (APA) Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations.

American Psychologist, 54, 594-604. (Reprint available through the APA Home Page: http://www.apa.org/journals/amp/amp548594.html)

struct_mult.doc

Table 1

Heuristic Data for Canonical Correlation Example.

	Predictor Var.		Criterion Var.				
Case No.	SECURE	PREOCC	NEURO	EXTRA	OPEN	AGREE	CONSC
1	35(-1.593)	52(.747)	51	46	57	51	50
2	36(-1.473)	55(1.268)	53	51	45	52	45
3	45(-.386)	51(.573)	48	35	56	55	48
4	45(-.386)	49(.226)	54	42	51	38	63
5	48(-.024)	54(1.095)	50	43	38	63	37
6	50(.217)	39(-1.512)	46	50	39	49	39
7	52(.459)	44(-.643)	52	48	57	46	43
8	53(.579)	39(-1.512)	39	59	50	53	58
9	59(1.304)	49(.226)	48	62	43	34	55
10	59(1.304)	45(-.469)	45	50	47	59	53

Note. Values in parentheses for SECURE and PREOCC are Z scores.

Table 2

Standardized Canonical Function Coefficients and Structure Coefficients for Function One.

Variable	Funct. Coef.	r_s	r_s^2
SECURE	.284	.752	56.55%
PREOCC	-.809	-.973	94.67%
NEURO	**-1.217**	**-.860**	73.96%
EXTRA	-.163	**.567**	32.15%
OPEN	.064	-.120	1.44%
AGREE	**-.745**	-.165	2.72%
CONSC	-.367	.190	3.61%

Note. r_s = structure coefficient. r_s^2 = squared structure coefficient times 100. The largest two function and structure coefficients for the criterion variables are in bold.

struct_mult.doc

Table 3

Calculation of the Synthetic Predictor Variable.

SECURE	(WEIGHT)	+	PREOCC	(WEIGHT)	=	Synthetic Predictor
-1.593	(.284)	+	.747	(-.809)	=	-1.057
-1.473	(.284)	+	1.268	(-.809)	=	-1.444
-.386	(.284)	+	.573	(-.809)	=	-.573
-.386	(.284)	+	.226	(-.809)	=	-.292
-.024	(.284)	+	1.095	(-.809)	=	-.893
.217	(.284)	+	-1.512	(-.809)	=	1.285
.459	(.284)	+	-.643	(-.809)	=	.651
.579	(.284)	+	-1.512	(-.809)	=	1.388
1.304	(.284)	+	.226	(-.809)	=	.188
1.304	(.284)	+	-.469	(-.809)	=	.750

Table 4

Standardized Weights and Structure Coefficients Throughout the General Linear Model.

Analysis	Standardized Weight Analog	Structure Coefficients Correlate:
Multiple regression	beta (β)	Observed predictors with synthetic $\hat{Y}$.
Canonical correlation analysis	canonical variate/function coefficients	Observed predictors/criterion variables with synthetic canonical variate/function.
Descriptive discriminant analysis	discriminant function coefficients	Observed dependent variables with synthetic discriminant function.
Factor analysis	pattern coefficients	Observed variable with synthetic/latent factor.
Confirmatory factor analysis	pattern coefficients	Observed variable with synthetic/latent factor.
Multivariate analysis of variance	No consensus on name/Same as discriminant function coeff.	Observed dependent variables with synthetic dependent variable.

struct_mult.doc

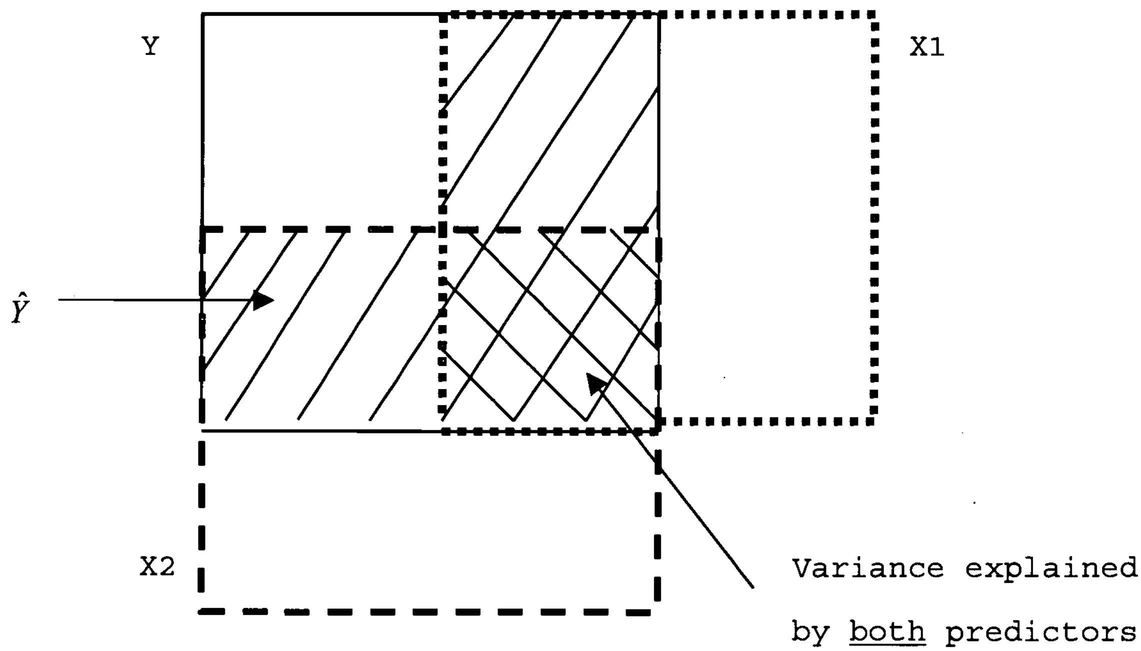

Figure 1. Venn diagram of multiple regression with two predictors for R^2 = .75 with multicollinearity.

struct_mult.doc

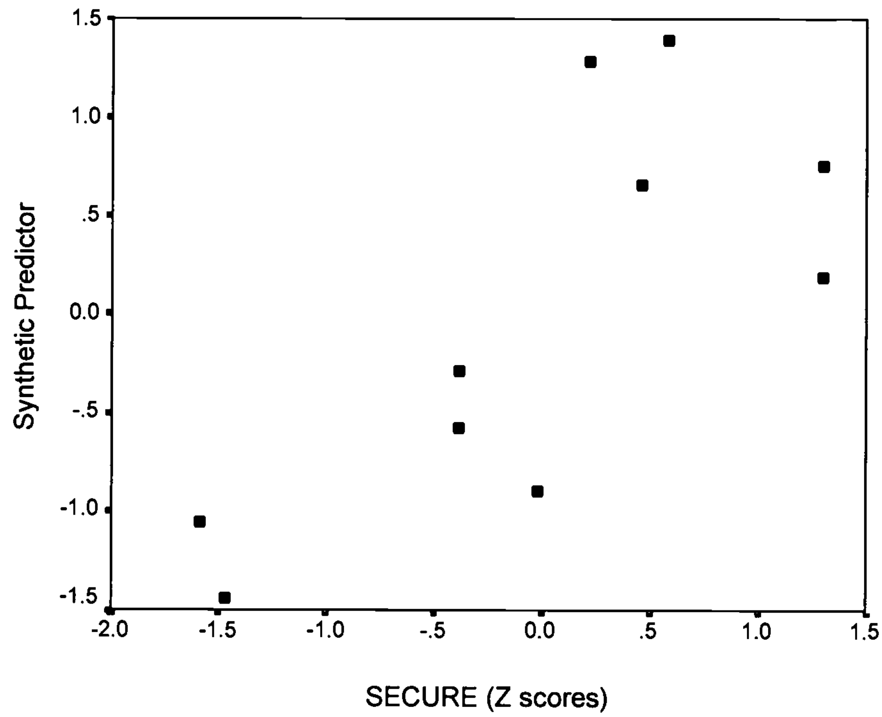

Figure 2. Scatterplot between the observed SECURE predictor and the synthetic predictor variable.

struct_mult.doc

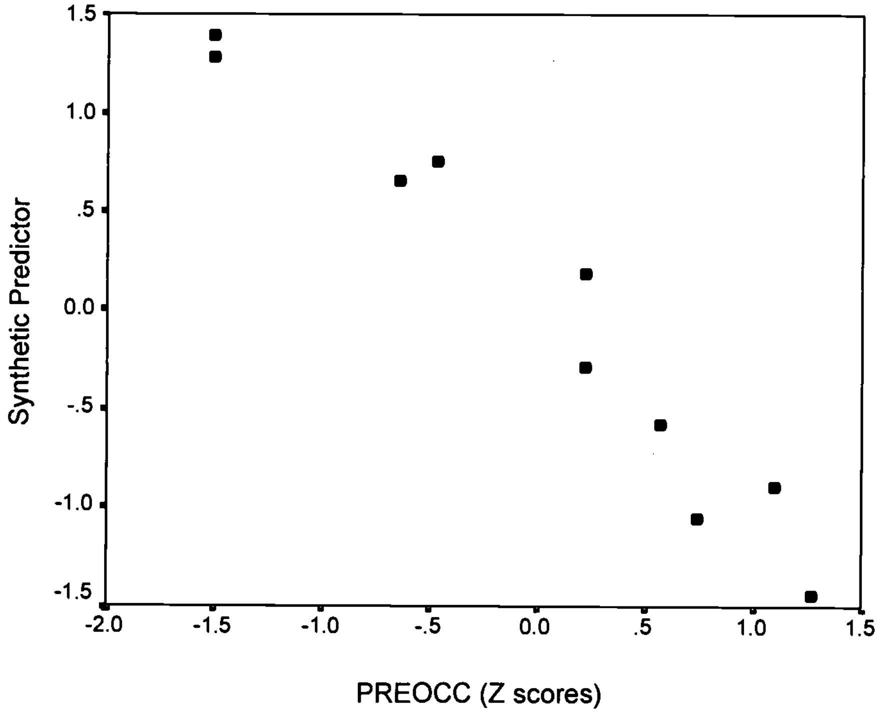

Figure 3. Scatterplot between the observed PREOCC predictor and the synthetic predictor variable.

struct_mult.doc

Appendix

```
title 'structure coefficients demo'.
comment Run canonical correlation.
MANOVA
  neuro extra open agree consc WITH secure preocc
  /PRINT=SIGNIF(MULTIV EIGEN DIMENR)
  /DISCRIM=STAN ESTIM COR ALPHA(.999).
```

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: The logic and interpretation of structure coefficients in multivariate general linear model analyses	
Author(s): Robin K. Henson	
Corporate Source: University of North Texas	Publication Date: April 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	**Level 2A**	**Level 2B**
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Robin K. Henson	Printed Name/Position/Title: Robin K. Henson/Assistant Professor	
Organization/Address: Dept. of Technology and Cognition, P.O. Box 311337, Denton, TX 76203-1337	Telephone: 940-369-8385	Fax: 940-565-2185
	E-mail Address: rhenson@unt.edu	Date: 6/20/02

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:	
ERIC Clearinghouse on Assessment and Evaluation **1129 Shriver Laboratory (Bldg 075)** **College Park, Maryland 20742**	**Telephone: 301-405-7449** **Toll Free: 800-464-3742** **Fax: 301-405-8134** **ericae@ericae.net** **http://ericae.net**

EFF-088 (Rev. 9/97)